国内外算法治理研究的主题比较与发展趋势 *

张 涛 王 铮
（黑龙江大学信息管理学院，哈尔滨 150080）

摘 要:［**目的 / 意义**］随着生成式人工智能的快速发展，算法治理问题成为社会各界关注的焦点。对国内外对算法治理研究主题进行分析与比较，能够更好地了解算法治理领域研究现状和国内外异同，进而为我国算法治理研究提供参考。［**方法 / 过程**］以中国知网和 Web of science 数据库中的期刊文献为数据来源，采用 LDA 主题模型对文献主题进行挖掘分析，得出国内外算法治理领域的主题热点及研究框架，并对国内外研究主题进行比较分析与趋势展望。［**结果 / 结论**］根据对算法治理研究主题的分析与比较，发现国内外研究既存在共性也存在差异。算法治理基础与法律规制是国内外共同关注的研究主题；在细化主题方面，国内侧重于市场监管和推荐算法的规制，而国外侧重于个人数据保护与算法技术层面研究。未来发展趋势主要聚焦于应用领域的算法治理政策法规研究、算法应用向善和服务从善研究、人工智能算法可解释性研究。

关键词: 算法治理 主题分析 LDA 模型 国内外比较
分类号: G353
DOI: 10.31193/SSAP.J.ISSN.2096-6695.2024.03.01

0 引 言

目前，人工智能技术快速发展，人工智能算法已经广泛应用于人类社会的各个方面，并在日常生活中发挥着重要而不可替代的作用。随着算法应用领域的不断扩展，算法偏见、算法黑箱、算法歧视等各种风险也开始出现，这可能会给社会稳定乃至国家安全带来一定程度的影响。特别是近年来以非法抓取使用个人信息与商业平台大数据杀熟为典型的算法推荐负面事件的频繁发生，进一步将以算法治理为深层次逻辑的平台治理推向网络社会治理的前台。这不仅涉及平台信息内容与平台经济驱动等关乎经济利益乃至网络权力问题，更关系到国家的信息主权安全。2021 年，国家互联网信息办公室、中央宣传部、教育部等联合发布《关于加强互联网信息服务算法综合治理的指导意见》，提出建立算法安全治理机制、构建算法安全监管体系、促进算法生态规范发展的指导意见，这反映出国家对算法治理问题的高度重视，算法治理已经成为国家安全治理的重要方面。2021 年 12 月，国家互联网信息办公室、工业和信息化部、公安部等联合发布《互

* 本文系国家社会科学基金一般项目“数智环境下情报分析算法风险治理路径研究”（项目编号：22BTQ064）的研究成果之一。
［作者简介］张涛（ORCID：0000-0002-3367-4541），男，教授，博士，研究方向为政策文本计算、数据与算法安全治理，Email：zhangtao@hlju.edu.cn；王铮（ORCID：0009-0007-3200-913X），男，硕士生，研究方向为算法治理，Email：1027590326@qq.com。

联网信息服务算法推荐管理规定》，明确算法推荐服务提供者的主体责任，着力提升防范化解算法推荐安全风险的能力。随着生成式人工智能软件 ChatGPT 与 Sora 的横空出世，各界掀起了人工智能内容生成算法的应用热潮，但随之产生的虚假新闻、算法歧视、信息泄露的风险问题也限制了生成式人工智能算法的健康发展。此后，国家互联网信息办公室联合六部委快速响应，于 2023 年 7 月正式颁布了《生成式人工智能服务管理暂行办法》，在第一章第四条对提供和使用生成式人工智能服务提出了规定，要求“采取有效措施防止产生各种歧视，不得实施垄断和不正当竞争行为”，促进生成式人工智能的科学发展和规范应用。

由此可见，人工智能算法治理已经引起当前社会的广泛关注及国家相关部门的高度重视。在学术研究领域，如何规避或消除算法风险，使算法运行在阳光之下，从而有效推动科技发展与社会进步，也成为国内外算法治理领域的研究热点。

1 相关研究

对国内外算法治理领域的研究进展和关注热点进行综合分析，有助于全面掌握该领域的整体发展与研究状况。在国内算法治理领域，已经有学者对研究主题及演进趋势进行了研究。例如：彭茜[1]采用文献计量分析方法与 Citespace 软件对算法治理领域文献进行关键词共现与突变检测研究，得出了国内算法治理领域的研究热点及演进趋势，并提出算法治理研究未来的发展方向；何美等[2]基于知识图谱理论和 Citespace 软件对国内新闻算法研究文献，从发文作者、期刊、机构等多个角度进行文献计量分析，发现新闻算法研究的研究主题包括算法应用、算法变革与算法治理，提出该领域研究可以分为兴起、升温和爆发三个发展阶段；邝岩等[3]利用 Citespace 软件对算法治理研究的学科、关键词、时间序列进行分析，进而梳理出算法治理的演进脉络，并构建了包含算法技术、应用场景、运转特征、风险问题、路径构建五个研究维度的算法治理理论体系；张涛等[4]选取我国算法治理领域的政策文本与科研论文进行主题识别和相似度计算，得出我国算法治理政策与科研的协同情况；黄萃等[5]基于学科交叉视角，以国内外人工智能治理论文为研究对象，采用关键词共现、突现检测、网络分析等方法，从总体发展、交叉广度与跨学科融合等方面揭示国内外人工智能治理研究的特征与差异。

国外对算法治理领域文献主题进行综合研究的论文相对较少，多数学者聚焦于某领域中的某些具体问题开展研究。例如：Latzer 等[6]通过一个理论模型来衡量算法治理的重要性，并采用经验混合方法来测试其在不同的生活领域所展现的不同之处；Issar 等[7]为算法治理领域提供了一个研究框架，认为算法治理问题受到普遍关注的三个方面包括算法权力、算法歧视和算法识别；Srivastava[8]针对科技公司证明算法治理能够在一定程度上控制信息污染、信息偏见与信息歧视等相关算法危害，建议更多学者开展这个领域的研究；Martin 等[9]对公司在何种情况下使用算法进行业务决策具有合法性进行了探讨，为企业算法决策提供了新视角；Basukie 等[10]探讨了共享经济平台数据治理和算法管理问题，发现算法对共享经济平台具有负面影响，并建议将法律和道德作为新兴市场的主要监管手段。

综上所述，当前对算法治理领域的研究进展和主题分析多采用传统文献计量方法开展研究，尚

缺少从语义模型视角对国内外算法治理文献进行比较分析的系统研究。鉴于此，本文以中国知网和Web of science 数据库中的期刊论文为数据来源，采用 LDA（Latent Dirichlet Allocation）模型对相关文献主题进行数据挖掘，旨在通过对国内外文献的算法治理研究主题进行可视化呈现与比较分析，进一步发现国内外研究存在的共性与差异性，并展望其发展趋势，以便从中得到启示与借鉴。

2 数据来源与研究方法

2.1 数据来源

本文国内算法治理研究文献来源于中国知网的期刊论文数据库平台，选取南京大学 CSSCI 和北京大学《中文核心期刊要目总览》收录期刊的论文并集作为国内文献研究样本。国外算法治理研究文献选取 Web of Science 数据库中的核心合集；为便于统计分析，将国内学者发表在国外期刊的英文论文按照国外文献计算。在数据库检索中，选择“算法治理（algorithmic governance）”、“算法规制（algorithmic regulation）”和“算法权力（algorithmic power）”为中英文主题词，文献类型选择“论文”，不限定发表年份。共检索到中文文献 590 篇，英文文献 2 592 篇。经过筛选，去除主题不相关文献，得到中文文献 415 篇，英文文献 358 篇。最后，将标题、摘要、关键词字段导出纯文本文件作为研究数据样本。

2.2 研究方法

研究方法选择文献主题识别与分析方法，主要采用自然语言处理的主题挖掘典型模型 LDA 及工具软件 Python。LDA 模型能够从大量的文本语料中挖掘出潜在的主题结构，在使用文献摘要作为语料构建语料库时抽取主题词的准确度较高，且主题中的语义信息较为清晰[11]。研究流程主要包括数据收集、数据预处理、主题识别、主题共现和比较分析 5 个步骤。①在中国知网和 Web of Science 数据库中检索所需文献数据并下载，作为本文的数据来源；②利用 Jieba 对文献数据进行分词处理，通过提取文献数据中的关键词形成关键词表，在文献数据进行挖掘时将无实际意义及影响模型效果的词进行过滤，将其与哈工大停用词表①合并成为无效词表；③使用“困惑度”确定最优主题数目，通过 LDA 主题模型将文献数据进行聚类，得到“主题—词项”表，进行主题识别；④使用 Python 生成主题词共现矩阵，根据结果生成主题词共现矩阵形成框架，将主题词共现矩阵导入 Ucinet 软件中保存为 Pajek 类型文件，之后将文件导入 VOSviewer 分析软件中生成主题词共现图谱[12]；⑤对国内外相关研究主题及框架进行比较分析。具体研究步骤如图 1 所示。

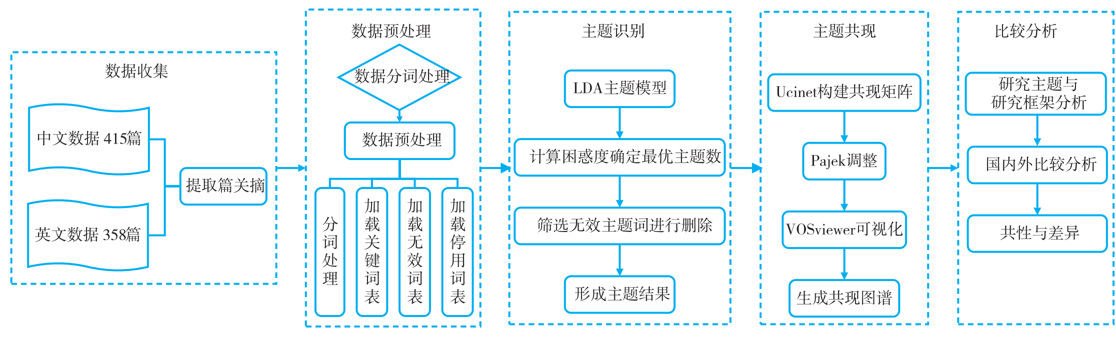


图 1　研究步骤

3　国内算法治理文献主题分析

通过对国内算法治理相关文献进行数据挖掘与主题分析，我们得出国内算法治理领域的研究热点和主题分布，如表 1 所示。

表 1　国内算法治理研究主题分布

	主题标识	词项（与主题相关的前 6 个高概率词）
主题 1	算法风险	数据、算法权力、法律、算法解释权、算法媒体、决策
主题 2	法律规制	人工智能、监管、政策、金融、法律、算法伦理
主题 3	政治应用	用户、政府、数据、人工智能、媒体、新闻传播
主题 4	市场监管	法律、数据、消费者、人工智能、定价算法、市场
主题 5	算法治理	自我治理、数据、框架、黑箱、数据保护、人工智能
主题 6	信息保护	透明度、信息茧房、用户、数据、个性化、个人信息保护法
主题 7	算法竞争	算法共谋、政治、算法政治、反垄断、垄断协议、监管
主题 8	推荐算法	算法歧视、信息不对称、数据、技术中立、歧视、算法推荐
主题 9	金融领域	垄断规制、金融科技、数字经济、金融消费者、规制路径、优势

在表 1 所示的国内算法治理“主题—词项”基础上，绘制主题词共现图谱，如图 2 所示。在共现图谱中，主题词圆圈大小代表主题词重要程度，连线表示主题词之间的相互联系。

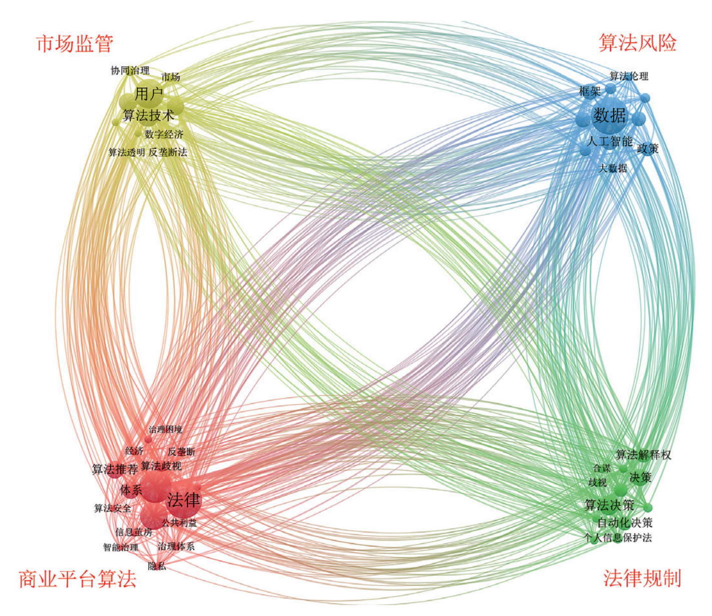


图 2　国内算法治理主题词共现图谱

从表 1 和图 2 可知，通过对国内算法治理领域文献的主题词共现图谱进行分析，可以将国内算法治理领域的研究热点和主题归纳为 4 个方面，包括算法风险、商业平台算法、市场监管和法

律规制。由此，可以形成国内算法治理领域的大致研究框架，包括事前、事中、事后三个环节的算法治理。“算法风险”属于事前环节，研究侧重于在风险产生之前进行防范与规避；“商业平台算法”“市场监管”属于事中环节，研究侧重于对目前所应用算法进行监管与改进；“法律规制”属于事后环节，研究侧重于对已经形成的算法风险进行政策法律层面的规制。最终形成围绕算法风险与平台算法问题，对人工智能算法进行事前风险防范、事中市场监管与事后法律规制的研究框架，如图3所示。

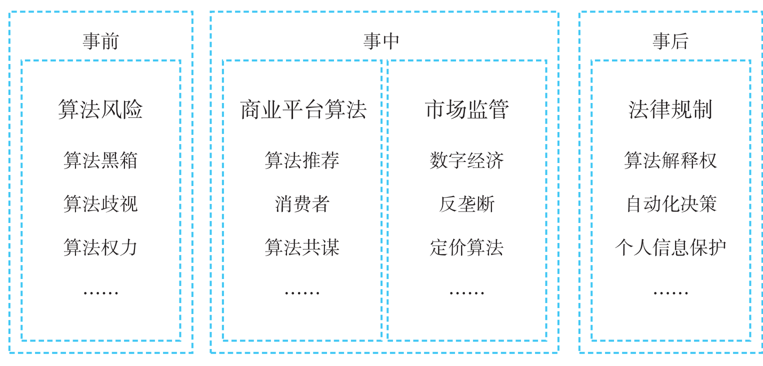


图3　国内算法治理的研究框架

3.1　算法风险

该主题涉及的主题词包括“算法黑箱”“算法歧视”“算法权力”等，这些均为算法在应用过程中常见的算法风险。此外，“信息茧房”“算法伦理”等也是较为常见的算法风险类型。“算法权力”是掌握算法技术的个人和企业利用自身的技术优势和行业便利，把控社会资源及信息，引导政府做出决策，从而形成的一股不可小觑的力量。在算法权力形成之后，企业就能够对公共资源或政策产生一定程度的干涉，这对于维护社会公平和稳定会有一定的负面影响。国内学者通过对算法风险进行研究，为我国算法治理和规制提供理论依据与对策建议。例如：张凌寒[13]探究了算法权力的兴起及其基础，指出了算法权力在商业领域和公权力领域产生异化的风险，包括消费者权益受损、公权力运行失范等问题，最后提出算法权力规制的基本思路与建议；汝绪华[14]重点对算法与政治的融合进行了探究，描述了算法政治的风险及其发生逻辑，为规避算法政治风险提出了建议，包括建立行业规范、优化算法设计、提升公众算法素养等；张涛等[15]基于风险社会理论、监管沙盒理论构建算法识别模型，对智能情报分析项目中数据与算法风险进行识别并验证其有效性，提出防范与化解重大安全风险的对策建议。

3.2　商业平台算法

该主题涉及的主题词包括“算法推荐”“消费者”“算法共谋”等，说明平台利用算法“精准算计消费者”的商业行为备受关注。用户自主搜索或被动获取的信息，多是由算法根据用户以往的搜索记录或观看历史进行推荐。这种方式能够有效提高用户检索和获取信息的效率，使用户更加方便地获取自己所需信息。但是，在这种信息获取过程中，用户所看到的都是“推荐算法想让你看到的”。例如，短视频平台中所浏览的内容大都是通过推荐算法根据用户喜好进行筛选所呈现，致使用户只接收和选择愉悦自己的内容，这种现象被称为“信息茧房”[16]。推荐算法需

要获取用户的个人信息来绘制用户画像，然后根据所得出的结果进行推荐，在这个过程中用户数据的用途和去向往往是不为人知的，这就造成了个人隐私信息泄露的风险。由此可见，算法在生活中的作用日益明显，每个社会中的个体都会或多或少与算法产生直接或间接的联系。为保证信息流动的通畅性以及个人信息不被随意泄露和滥用，目前已经形成基于算法应用场景的研究热点。邓胜利等[17]基于扎根分析理论，探索算法推荐服务风险下用户的应对行为，并构建用户行为模型，提出从算法素养、学习行为与风险参与三个方面降低算法推荐服务风险的策略。周颖玉等[18]基于对算法推荐伦理失范风险的分析，从个人观念、科技本身、财富关联、规约机制审视并探讨算法伦理失范的原因，提出从伦理、法律、技术和文化等维度建构人机和谐生态以规制算法伦理行为。

3.3 市场监管

该主题涉及的主题词包括“数字经济”“反垄断”“定价算法”等。算法对于企业而言是一种技术经济元素，通过对算法应用来挖掘信息中蕴含的经济价值，可以提升企业的交易效率与竞争优势，从而有效推动数字经济的发展。但是，近年来大数据杀熟、价格歧视等现象层出不穷。一些商业平台作为数据集中和中转的枢纽，利用其平台优势来掌控数据，并制定掠夺性价格。这就侵犯了消费者的权益，扰乱了市场秩序，形成不正当竞争甚至行业垄断的现象。朱虹影[19]以反垄断的视角对互联网算法共谋产生背景、治理困境进行研究，提出明确责任主体、建立事前监督机制等建议。叶明等[20]对数字经济时代算法价格歧视行为的风险进行了探讨，阐述了算法价格歧视所存在的风险，并建议拓宽《反垄断法》的主体范围，明确责任主体，有效规制算法价格歧视行为。

3.4 法律规制

该主题涉及的主题词包括“算法解释权”“自动化决策”“个人信息保护”等。目前，算法在社会各方面的应用都较为广泛，尤其在教育、医疗、资源分配、公共决策等方面发挥着重要的作用，有效提高了社会运行的效率。但是，在应用过程中也产生各种风险问题，形成制约算法健康发展的瓶颈。因此，对上文所提到的算法风险进行有效的法律规制，成为算法治理领域中的研究热点。我国的《个人信息保护法》首先确立了算法自动化决策治理的基本框架，此后《互联网信息服务算法推荐管理规定》进一步规范了互联网信息服务算法推荐活动。丁晓东[21]认为，算法的法律规制应该根据不同的算法应用场景选择不同的规制方式，构建算法公开、数据赋权与反算法歧视的制度。郑戈[22]在法学视角下对如何用法律规制算法和算法强化法律进行研究，并对未来法律规制算法进行了展望。张涛等[23]对我国算法治理政策法规文本进行编码分析并提取出我国算法治理政策法规框架，总结我国算法治理存在的问题并提出相关建议。此外，国内相关学者还针对单方面的算法风险进行法律规制研究，例如，蒋慧等[24]对电商平台个性化推荐算法的风险、成因、实践中的困境和应对策略进行了较为全面的研究，旨在推进平台算法规制制度的完善。

4 国外算法治理文献主题分析

采用和上文同样的研究方法，我们对国外算法治理相关文献进行数据处理与主题分析，得出国外算法治理领域的研究热点和主题分布，如表 2 所示。

表 2　国外算法治理研究主题分布

	主题标识	词项（与主题相关的前 6 个高概率词）
主题 1	法律治理	artificial intelligence（人工智能）、law（法律）、decision（决策）、legal（法律）、algorithms（算法）、fairness（公平）
主题 2	社交媒体	social（社会）、media（媒体）、social media（社交媒体）、experience（经验）、personalization（个性化）、citizens（公民）
主题 3	算法歧视	artificial intelligence（人工智能）、data（数据）、discrimination（歧视）、bias（偏见）、algorithms（算法）、social（社会）
主题 4	政治应用	data（数据）、algorithms（算法）、authors（作者）、regulation（监管）、politics（政治）、surveillance（监控）
主题 5	算法可解释性	artificial intelligence（人工智能）、data（数据）、algorithm（算法）、paper（论文）、algorithms（算法）、interpretability（可解释性）
主题 6	数据安全	data（数据）、protection（保护）、data protection（数据保护）、regulation（监管）、personal（个人）、GDPR（通用数据保护条例）
主题 7	算法监管	data（数据）、algorithms（算法）、machine learning（机器学习）、artificial intelligence（人工智能）、accountability（问责）、regulation（监管）
主题 8	市场应用	performance（表现）、market（市场）、data（数据）、continues（持续）、pricing（定价）、hidden（隐藏）
主题 9	平台经济	workers（工作者）、platform（平台）、platform work（平台工作）、management（管理）、algorithmic management（算法管理）、labor（劳工）

在表 2 的基础上，绘制国外研究主题词共现图谱，如图 4 所示。在共现图谱中，主题词圆圈大小代表主题词重要程度，连线表示主题词之间的相互联系。

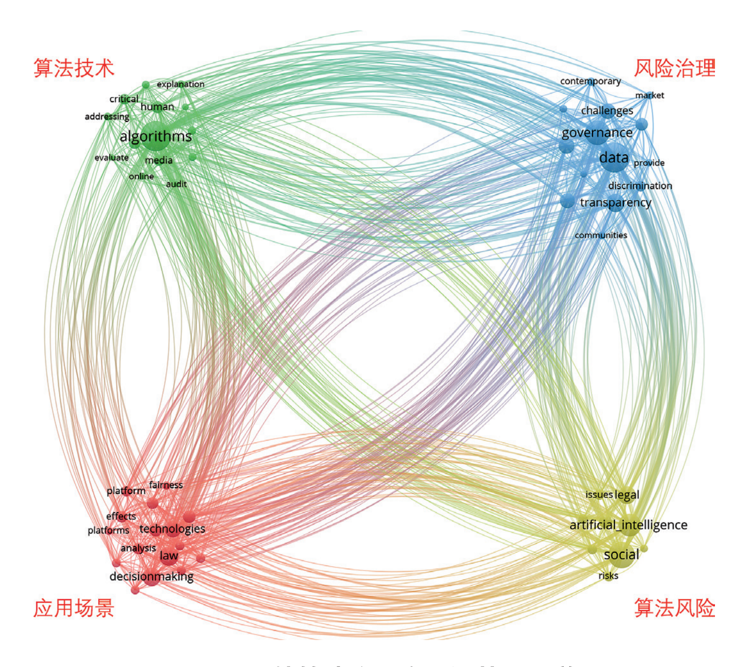


图 4　国外算法治理主题词共现图谱

从表 2 和图 4 可知，通过对国外算法治理领域文献的主题词共现图谱进行分析，可以发现国外算法治理领域的研究热点和主题包括以下 4 个方面：算法技术；应用场景；算法风险；风险治理。从这 4 个方面可以大致分析出，国外算法治理领域主要以算法技术为支撑，以应用场景为基础对算法风险及其治理进行研究，从而形成由技术到应用、由风险到治理较为全面的研究框架。如图 5 所示。

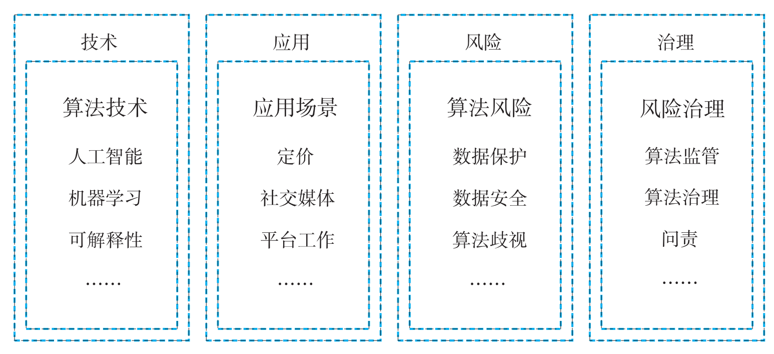


图 5　国外算法治理的研究框架

4.1　算法技术

“算法技术”方面的主题词主要有“人工智能”“机器学习”“可解释性”等。算法技术主要指机器学习和深度学习中的具体算法，如随机森林、朴素贝叶斯、神经网络、计算机视觉等，而算法可解释性指的是对于具体的人工智能算法行为做出能够为人们所理解的解释。随着人工智能技术的发展，算法已经应用于计算机视觉、自然语言处理、语音识别等实用性领域，并逐渐超越了人类的工作效率。但是，算法运作的过程像是一个“黑盒”，尤其是深度神经网络，只需要进行输入并运行就能得到一个决策结果。同时，该“黑盒”进行决策的依据和具体过程却是不为人知的，对于一些比较重要的工作，没有得到验证的决策结果不能够轻易应用。例如，利用病人数据进行辅助医疗决策，如果不能明确算法进行决策的过程和依据，将会造成严重的后果。人工智能算法这种不可解释性对于算法的发展和应用来说是一种巨大的阻碍，该领域的学者们需要针对算法技术进行大量的研究与探索。例如：Kelly 等[25]探讨了人工智能在医疗保健领域进行大规模应用的主要挑战和局限性，包括技术限制、应用方法以及社会文化接受程度，并提出了相关的建议；Selbst 等[26]探讨了机器学习与其他制定决策规则的方式的区别，以及这些区别带来的解释性问题，并主张建立其他规范性评价机制，寻求对模型开发过程的解释。Lundberg 等[27]针对基于树的机器学习模型，如随机森林、决策树和梯度提升树，提出了一种树的解释方法，能够计算单个预测的最佳局部解释。

4.2　应用场景

“应用场景”方面的主题词主要包括“定价”“社交媒体”“平台工作”等。近年来，在大数据、云计算、物联网等技术的支持下，人工智能算法得到了迅速的发展，在自动驾驶、人脸识别、医疗辅助决策等方面的应用也越来越普遍。算法应用于社会生活中的各种具体场景，在不同

的场景中会产生不同的问题，因而对算法治理的研究也应结合具体的应用场景[3]。在众多算法应用场景中，社交媒体、网络平台、自动决策等是较为常见的领域，针对这些领域的算法治理内容包括推荐算法与个人信息保护。为保证算法应用过程中信息流动的通畅性，避免个人信息被随意泄露和滥用等方面的问题，目前已经形成了算法应用场景方面的研究热点。Butler 等[28]总结了分子和材料科学机器学习的最新进展，并概述了适合解决该领域研究问题的机器学习技术，以及该领域的未来方向。Bahrammirzaee[29]对金融市场中三种著名的人工智能技术，即人工神经网络、专家系统和混合智能系统进行了比较研究综述，发现在处理金融问题方面，非线性模式的精度优于传统统计方法。

4.3　算法风险

“算法风险”方面的主题词主要包括“算法歧视”“数据保护”“数据安全”等，具体体现在政治和商业两个领域。

在政治领域中，将算法应用于教育、医疗、司法等公共领域，能够极大地降低政府运行所需要的成本。原本需要多方面专家进行决策才能决定的事务，只需要使用算法和很少的人工干预就能完成，包括基于自主决策系统的辅助政治决策与基于算法的政治传播。但是，在算法与政治进行融合的过程中，产生算法风险的后果相较于其他领域来说更加严重，因此外国学者对算法政治风险展开了研究。例如，Coglianese 等[30]讨论了政府机构使用人工智能算法进行行政决策的合理性，认为如果政府机构能够正确理解算法技术，那么对算法技术的应用就能够很好地符合传统法律规范，但是需要监管保障措施来使其决策更有效率。

在商业领域中，虽然算法具有中立性，但是在算法开发过程中，由于开发者具有自己的立场和观点，甚至存在偏见和歧视，会导致在算法的编写过程中加入主观性的偏向，或者使用的数据具有地域性与群体性。最终，都会导致运用到实际中的算法存在各种歧视的倾向，并且由于责任主体的不明确，难以进行追责。Rosenblat 等[31]通过对“优步”车主进行跟踪实证研究，发现“优步”通过信息权利不对称及算法修饰对合作车主的工作方式进行间接控制，因此他呼吁学者们更加关注平台的去中介化，在雇主和工人之间建立平等的权利关系。

4.4　风险治理

“风险治理”方面的主题词主要包括“算法监管”“算法治理”“算法问责”等。目前，针对算法在应用过程中产生的潜在风险治理问题，该领域专家学者已经产生出比较多的研究成果。算法可解释性是对于算法本身进行的治理，对存在偏向和误差的算法及时进行调试和优化，有助于提高算法的准确性，避免潜在风险造成的危害。同时，需要国家出台相关的政策法规来对算法进行有效的监管和审核，使算法运行在法律的框架之内。美国在 2022 年出台了《算法责任法案》，该法案要求自动化决策系统具有新的透明度和问责制，确保算法能够透明地运行，明确算法的责任主体。Goodman 等[32]总结了欧盟《通用数据保护条例》的出台和生效对机器学习算法的使用所产生的潜在影响，并认为这项法律是保障计算机科学专家带头设计算法及其评估框架的好机会。

随着智能设备的普及率不断提高，人们接触和使用到的算法也越来越多。但是，大部分人对算法了解不多甚至一无所知，这对识别算法和规避算法风险极为不利，因此提高人们的算法素养

也是算法治理的重要一环。Joëlle[33]对 22 名年轻人进行深入访谈，探讨了当代年轻人如何理解、感受和参与社交媒体上的新闻算法，以及这些经历如何有助于他们自身算法素养的提升。

5　国内外算法治理研究主题比较

通过对国内外算法治理研究文献进行主题挖掘，可以得出该领域在国内外的发展现状与研究热点。在此基础上，比较分析这些文献主题与研究热点，能够发现国内外算法治理研究中存在的共性与差异。

5.1　国内外研究共性

从整体来看，算法治理研究虽然出现较早，但在近几年才开始快速发展。国外研究在 2016 年美国发布《国家人工智能研究和发展战略计划》后进入快速发展期，国内研究则在 2017 年《新一代人工智能发展规划》发布后得到了学者们的重视。总体而言，目前该领域仍处于发展初期，研究成果相对于发展成熟的研究领域来说数量较少，在未来该领域还将会得到更大的发展。

算法风险作为算法治理领域的基础性研究，国内外都已经产出了大量的研究成果，如算法共谋、算法黑箱、算法歧视与偏见等。算法风险研究是进行算法治理的基础，只有明确算法存在哪些风险才能更好地针对风险进行治理和规制，因此国内外学者对算法风险研究都较为重视。

算法治理已经成为世界性问题，算法治理体系与治理能力不仅成为各国国家治理水平的象征，同时也是国家顺应时代发展与科技进步的内在需求。因此，在法律规制方面，国内外进行算法治理的方式也存在着众多共同点，例如，通过立法手段明确算法责任主体，通过分析算法风险的脉络提出治理框架，等等。在立法层面，国内外都制定了相关的法律为算法治理提供法律依据。例如：2021 年我国出台的《个人信息保护法》包含了自动决策算法的使用规范；美国在 2022 年出台的《算法责任法案》为算法带来监督和透明度；2024 年欧盟议会通过的《人工智能法案》包含了算法风险监管方法。

5.2　国内外研究差异

国内外研究差异体现在侧重点有所不同。国外算法治理在学科分布和学科知识来源方面更为均衡，涉及人文科学、社会科学、计算机科学等多个学科大类，在技术层面和社会层面有更多的跨学科知识流动，学科之间的交叉融合度更高。国内主要是社会科学领域的学者在关注算法治理问题的研究，对算法技术本身的研究相对较少，且多在某一特定领域内进行研究，缺乏不同学科之间的交叉融合，社会科学与其他学科之间的交叉与融合有待进一步提升。

国内算法治理的研究与实践侧重于市场监管和推荐算法的规制，包括对企业的市场竞争行为和推荐算法进行监管，监管内容涉及垄断行为、算法共谋、大数据杀熟等一系列问题。2021 年 12 月发布的《互联网信息服务算法推荐管理规定》，将推荐算法列为重点监管对象，针对推荐算法的安全风险评估等规制手段以及推荐算法带来的信息茧房、算法歧视等损害用户权益的问题进行监管，明确算法责任主体。实际上，在约束市场主体算法行为的同时，还需要对公共主体的监管，明确公共主体的算法责任。国家机关在实现行政自动化的过程中，也容易出现算法歧视、算法黑箱、结果失控的问题。

相较于国内来说，国外侧重于个人数据保护与算法技术层面的研究探索。2021 年欧盟出台《数据法案》，要求在用户使用由算法操控内容展示的情况下需要明确告知个人数据会被如何处理，同时保证用户退出的权利。对个人数据进行有效的保护在一定程度上能够规范算法的应用，使企业承担起社会责任，规范竞争行为。在算法技术方面，国外学者针对算法可解释性进行了众多研究，例如，Carvalho 等[34]对算法可解释性方面的文献进行了总结，并提出算法可解释性的方法和评估指标。算法运行过程的不透明性是阻碍算法技术进一步发展的重要因素，因此对算法进行解释能够使算法更加可信，更好地应用于专业领域并推动社会发展。

6 国内外算法治理研究发展趋势

从国内外算法治理研究的主题分析及比较结果来看，未来国内外算法治理领域研究的发展趋势主要体现在以下三个方面。

一是细化应用领域的算法治理政策法规研究。国内算法治理的研究主题更多地围绕“市场监管”“法律规制”“商业平台”等，而国外文献则聚焦于“算法监管”“算法问责”“风险治理”等主题词。从我国《互联网信息服务算法推荐管理规定》、美国《算法责任法案》、欧盟《人工智能法案》等代表性法案的颁布来看，国内外的算法治理呈现出自上而下的发展趋势，未来将会从顶层设计展开研究，聚焦于更加细化应用领域的政策法规，如医疗领域、自动驾驶领域、司法领域等。

二是加强算法应用“向善”和服务“从善”的研究。从国内外算法治理研究的共性可知，在算法应用与服务方面，算法垄断、算法共谋、大数据杀熟等问题较为突出。未来算法风险后果会严重影响社会稳定及国家安全，会有更多算法研究应用于政府决策中，因此引导算法应用“向善”将是国内外算法研究的重点。从国内外文献所包含的“反垄断”“定价算法”“消费者”等主题词来看，商业平台可能会利用其优势掌控数据，促使不正当竞争甚至行业垄断现象的蔓延，这就需要未来研究更多地关注算法服务“从善”。

三是注重人工智能算法可解释性的研究。在国外算法治理文献中出现了“算法黑箱”“算法偏见”“算法歧视”“可解释性”等主题词。其中，算法黑箱是算法偏见、算法歧视产生的根源，而人工智能算法的可解释性研究是打破算法黑箱的重要途径，也是当前算法治理事前风险防范与化解环节的关键点。因此，对人工智能算法可解释性研究可能在国内外都是具有前瞻性的重要发展趋势。

7 结 论

本文通过 LDA 主题模型抽取了国内外算法治理领域的核心期刊文献研究主题，并从中得出该领域的热点主题及研究框架。在比较分析国内外研究主题的基础上，总结出国内外算法治理研究的共性与差异，并从三个方面展望未来发展趋势。

总体来看，算法风险作为算法治理领域的基础性研究，国内外学者都较为重视并产出了大量

的研究成果；在法律规制方面，国内外算法治理方式也存在诸多共同点。在差异性方面，国外学者主要从算法风险、算法技术、法律规制和应用场景等方面进行研究，主题分布较为均匀，有着稳定的研究结构；国内研究侧重基于算法的市场竞争，着重对垄断规制与不正当竞争行为等问题进行探讨。另一方面，国外学者注重理论研究，为算法治理寻找具有普适性的治理方法和路径；国内则更注重于针对某一领域的特定算法问题进行研究，相较于国外更具有实践性，体现出自身的社会文化特点。

本文尚存在一定的局限性。一是数据来源仅限于核心期刊论文，文献样本虽具有代表性但也存在片面性。二是由于算法治理研究领域的中外文献所涉及的主题词极为宽泛且发展迅速，在文献采集中难免存在时间滞后或部分遗漏，可能会导致结论分析不够全面。在未来的研究中，我们将拓宽来源数据的采集范围，尝试从主题演化、主题扩散、时空分布等多角度进行全方位综合分析。

【注释】

①哈尔滨工业大学停用词表［EB/OL］.［2024-08-31］.https://github.com/gogo456/stopwords/blob/master/hit_stopwords.txt.

【参考文献】

［1］彭茜. 国内算法治理研究热点及演变趋势——基于CiteSpace软件的可视化分析［J］. 宁夏党校学报，2023，25（2）：88-95.

［2］何美，郑勇华. 基于CiteSpace的国内新闻算法研究的知识图谱分析［J］. 东南传播，2022（9）：56-60.

［3］邝岩，许晓东. 算法治理研究述评：演进脉络分析与理论体系构建［J］. 情报杂志，2023，42（3）：158-166.

［4］张涛，王瀚功，崔文波. 我国算法治理政策与科研主题协同研究——基于LDA与Word2vec融合模型［J］. 网络安全与数据治理，2023，42（8）：13-20.

［5］黄萃，黄施旗，付慧真. 学科交叉视角下人工智能治理领域知识流动与研究主题的国际比较研究［J］. 信息资源管理学报，2022，12（6）：98-110.

［6］Latzer M, Festic N. A guideline for understanding and measuring algorithmic governance in everyday life［J］. Internet Policy Review, 2019, 8(2): 1-19.

［7］Issar S, Aneesh A.What is algorithmic governance?［J］. Sociology Compass, 2021, 16(1): 1-14.

［8］Srivastava S. Algorithmic governance and the international politics of big tech［J］. Perspectives on Politics, 2023, 21(3): 989-1000.

［9］Martin K, Waldman A. Are algorithmic decisions legitimate? the effect of process and outcomes on perceptions of legitimacy of AI decisions［J］. Journal of Business Ethics, 2022, 183(3): 1-18.

［10］Basukie J, Wang Y, Li S. Big data governance and algorithmic management in sharing economy platforms: a case of ride sharing in emerging markets［J］. Technological Forecasting & Social Change, 2020, 161: 120310.

［11］祁颖，张涛. 国内外人文社科领域跨学科研究：文献主题对比与中国路径选择［J］. 情报科学. 2023，41（12）：81-90.

［12］Cui W, Li J, Zhang T, et al. A recognition method of measuring literature topic evolution paths based on K-means-NMF［J］. Knowledge Organization, 2023, 50(4): 257-271.

［13］张凌寒. 算法权力的兴起、异化及法律规制［J］. 法商研究，2019，36（4）：63-75.

［14］汝绪华. 算法政治：风险、发生逻辑与治理［J］. 厦门大学学报（哲学社会科学版），2018（6）：27-38.

［15］张涛，马海群. 智能情报分析中数据与算法风险识别模型构建研究［J］. 情报学报，2022，41（8）：832-844.

［16］彭兰. 导致信息茧房的多重因素及“破茧”路径［J］. 新闻界，2020（1）：30-38，73.

［17］邓胜利，段文豪，夏苏迪. PADM视角下算法推荐服务风险的用户应对行为研究［J］. 图书情报工作，2023，67（2）：14-22.

［18］周颖玉，柯平，刘海鸥. 面向算法推荐伦理失范的人机和谐生态建构研究［J］. 情报理论与实践，2022，45（10）：54-61.

［19］朱虹影. 反垄断视角下互联网平台算法共谋规制研究［J］. 法制博览，2023（15）：136-138.

［20］叶明，郭江兰. 数字经济时代算法价格歧视行为的法律规制［J］. 价格月刊，2020（3）：33-40.

［21］丁晓东. 论算法的法律规制［J］. 中国社会科学，2020（12）：138-159，203.

［22］郑戈. 算法的法律与法律的算法［J］. 中国法律评论，2018（2）：66-85.

［23］张涛，韦晓霞. 我国算法治理政策法规内容及框架分析［J］. 现代情报，2023，43（9）：98-110.

［24］蒋慧，徐浩宇. 电商平台个性化推荐算法规制的困境与出路［J］. 价格理论与实践，2022（12）：39-43.

［25］Kelly C J, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence.［J］. BMC Medicine, 2019, 17(1): 195.

［26］Selbst A D, Barocas S. The Intuitive appeal of explainable machines［J］. Foraham Law Review, 2018, 87(3): 1085-1139.

［27］Lundberg S M, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees［J］. Nature Machine Intelligence, 2020, 2(1): 56-67.

［28］Butler K T, Davies D W, Cartwright H, et al. Machine learning for molecular and materials science［J］. Nature, 2018, 559(7715): 547-555.

［29］Bahrammirzaee A. A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems［J］. Neural Computing and Applications, 2010, 19(8): 1165-1195.

［30］Coglianese C, Lehr D. Regulating by robot: administrative decision making in the Machine-Learning era［J］. Social Science Electronic Publishing, 2017, 105(5): 1147-1223.

［31］Rosenblat A, Stark L. Algorithmic labor and information asymmetries: a case study of Uber's drivers［J］. Social Science Electronic Publishing, 2015, 8583(4): 7.

［32］Goodman B, Flaxman S. European Union regulations on algorithmic decision making and a right to explanation［J］. AI Magazine, 2017, 38(3): 50-57.

［33］Joëlle S. Experiencing algorithms: how young people understand, feel about, and engage with algorithmic news selection on social Media［J］. Social Media + Society, 2021, 7(2).［2024-02-29］. https://doi.org/10.1177/20563051211008828.

［34］Carvalho D V, Pereira E M, Cardoso J S. Machine learning interpretability: a survey on methods and metrics［J］. Electronics, 2019, 8(8): 832.

Topic Comparison and Development Trend of Algorithm Governance Research at Home and Abroad

Zhang Tao　Wang Zheng

(School of Information Management, Heilongjiang University, Harbin 150080, China)

Abstract: [**Purpose/Significance**] With the rapid development of generative artificial intelligence, the problem of algorithm governance has become the focus of the industry. The research topic of algorithm governance is analyzed and compared at home and abroad, which can better understand the research status of algorithm governance and the differences at home and abroad, and then provide reference for the research of algorithm governance in China. [**Method/Process**] Taking the journal literature in CNKI and Web of science database as the data source, the LDA topic model is used to mine and analyze the literature topics, and the hot topics and research framework in the field of algorithm governance at home and abroad are obtained, and research topics at home and abroad are compared and analyzed, the trends of the research were also analyzed. [**Result/Conclusion**] According to the research topics and research frameworks of algorithm governance, it is found that there are both commonalities and differences in domestic and foreign research. Basic research and legal regulation of algorithm governance are the commonalities of domestic and foreign research, while there are some differences in the detailed topics at home and abroad. Domestic research focuses on market regulation and regulation of recommended algorithms, while foreign research focuses on personal data protection and algorithm technology. The development trend in the field of algorithm governance at home and abroad mainly focuses on the research of the refining the research of algorithm governance policies and regulations in the application field, strengthening the research of algorithm application to good and service from good, and explainability of artificial intelligence algorithms.

Keywords: Algorithm governance; Topic analysis; LDA model; Comparison at home and abroad

（本文责编：任全娥）